DOCUMENT RESUME

ED 309 189 TM 013 660

AUTHOR van der Linden, Wim J.
TITLE Applications of Decision Theory to Test-Based Decision Making. Project Psychometric Aspects of Item Banking No. 23. Research Report 87-9.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE Dec 87
NOTE 64p.
AVAILABLE FROM Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Classification; *Decision Making; Foreign Countries; Item Banks; Mastery Tests; Mathematical Models; Placement; Predictive Measurement; Problem Solving; Scores; Selection; Statistical Analysis; *Test Results; Test Theory
IDENTIFIERS *Decision Theory; Sequential Testing

ABSTRACT

The use of Bayesian decision theory to solve problems in test-based decision making is discussed. Four basic decision problems are distinguished: (1) selection; (2) mastery; (3) placement; and (4) classification, the situation where each treatment has its own criterion. Each type of decision can be identified as a specific configuration of one or more of the following elements: a test that provides the scores on which the decisions are based; one or more treatments with respect to which decisions are made; and one or more criteria by which the successes of treatments are measured. For each type of decision, further restrictions or generalizations may hold, such as multivariate test scores, sequential testing, multiple criteria, multiple populations, and quota restrictions. In some applications, combinations of the basic types of decisions may occur. Samples of decision problems illustrate the optimization of the Bayes utility for each possible decision. Examples are given for selection decisions with linear utility, mastery decisions with threshold utility, placement decisions with normal-ogive utility, classification decisions with threshold utility, and combinations of basic decisions. Nine figures illustrate decision systems, and one table gives data for an application. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED309189

Applications of Decision Theory to Test-based Decision Making

Research Report 87-9

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. NELISSEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Wim J. van der Linden

department of EDUCATION

University of Twente

Division of Educational Measurement and Data Analysis

TM013660

Project Psychometric Aspects of Item Banking No.23

Colofon:
Typing: Mevr. L.A.M. Padberg
Cover design: Audiovisuele Sectie TOLAB Toegepaste Onderwijskunde
Printed by: Centrale Reproductie-afdeling

Applications of Decision Theory to Test-based Decision Making

Wim J. van der Linden

Invited contribution to R.K. Hambleton & J.N. Zaal (Eds.), New developments in testing: Theory and applications. Amsterdam, The Netherlands: North-Holland.

Abstract

The use of Bayesian decision theory to solve problems in test-based decision making is discussed. Four basic decision problems are distinguished: selection, mastery, placement and classification. For each type of decision, further restrictions or generalizations may hold, namely multivariate test scores, sequential testing, multiple criteria, multiple populations or quota restrictions. In some applications, combinations of the basic types of decisions may occur.

Ha

Applications of Decision Theory to Test-based Decision Making

Historically, the use of tests has its roots in the necessity of selection and placement decisions in education, the army, and the public administration. This is excellently demonstrated in DuBois's (1970) historiography of such cases as Binet's pioneering work on developing a test for the assignment of retarded children to special education, the testing of conscripts for placement in the army during World War I, and the examination of applicants for the civil service in ancient China. It is no coincidence that in each of these fields decision making is characterized both by a high visibility and a massive number of examinees. In such cases it seems logical to use tests to base decisions on.

Inspired by these early successes, decision makers have been using tests ever since. Nowadays, the use of tests has pervaded such fields as the admission of students to schools, the selection of personnel in public as well as private settings, the assignment of clients to therapeutic treatments, the choice of careers in vocational guidance situations, pass-fail decisions in instructional units, certification, personnel review, tracking decisions in individualized study systems, and the evaluation of training programs. Although their contents and format may vary, it is inconceivable that the use of tests will ever leave these fields.

It is conspicuous that, although the practice of testing has its roots firmly in decision making, test theory has been developed mainly as a theory of measurements. This was already manifest in Spearman's pioneering work on what is now known as classical test theory. In this theory, test scores are modeled as a linear combination of a true score and an error of measurement and the concern is primarily in quantities such as the reliability coefficient, the standard error of measurement — as well as in their properties as a function of test length, item selection, and the like. Modern item response theory shows the same concern with measurement (ability estimation), and was not conceived as a theory of decision making either. The history of test theory shows a few exceptions, though, of which the publication of the Taylor–Russell (1939) tables, with their subsequent influence on the testing literature, and Cronbach and Gleser's (1965) Psychological tests and personnel decisions deserve special mention. To date, the latter has been the first and only monograph attempting to provide test–based decision making with a sound theoretical basis.

Recently, however, the situation has changed somewhat and some test theorists are now seriously involved in attempts at modeling and optimizing the use of test scores for decision making. A major impetus for this has come from the introduction of modern instructional systems as individualized instruction, learning for mastery, and computer–aided instruction. In such systems, testing

primarily serves instructional decision making, and an important task of their developers is to design optimal decision procedures. A seminal paper by Hambleton and Novick (1973) was the first to point at the challenge of these developments to classical test theory.

Although test theory has long ignored decision problems, at a more abstract level the study of optimal rules for decision making has had a long tradition in statistics dating back to early publications such as von Neumann and Morgenstern (1944) and Wald (1950). More up-to-date treatments of statistical decision theory can be found in DeGroot (1970) and Ferguson (1967); a short but excellent introduction is given in Lindgren (1976, chap. 8). It is the primary intention of this paper to demonstrate how the various types of decision making problems in testing can be solved using the framework of statistical decision theory. In particular, in doing so, an (empirical) Bayesian point of view will be assumed. In the next sections, first some decision-theoretic notions will be introduced. Then, a classification of all possible types of test-based decisions will be given. Subsequently, the main part of this paper showing how problems with respect to the various types of decisions can be solved using Bayesian decision theory is addressed. In the final section some results will be presented on the application of decision theory to the problem of simultaneous optimization of combinations of

different types of decisions in individualized instruction systems.

Elements of a Statistical Decision Problem

Statistical decision problems arise when a decision maker is faced with the need to choose a preferred action; the outcome of the action depends on the state of nature about which only partial information is available. A simple example is the vacationer who has to decide whether or not to go for the beach, but must rely on a forecast for information about the weather; a more sophisticated one is the researcher who has to decide on the basis of sample data which of his/her hypotheses holds for a given population.

The set of all possible states of nature relevant for the decision problem is known as the state space in decision theory. This set will be denoted by Ω whereas a numerical parameter ω will be used to index the 1 individual states in Ω (When Ω is discrete, ω_s, $s = 1, \ldots, S$, will be used.) Let A be the set of all possible actions from which the decision maker can choose. Technically, A is known as the action space. Individual actions will be denoted by a (continuous action space) or a_t, $t = 1, \ldots, T$, (discrete action space). For each action $a \in A$, the decision maker is confronted with certain consequences. These consequences depend not only on the action chosen but also on the (unknown) state $\omega \in \Omega$ nature is in. Some of these

consequences may be positive; others negative. It is supposed that the decision maker is able to summarize the consequences of his/her actions given the true state of nature into an evaluation on a numerical scale. As this scale is assumed to run from negative to positive evaluations, it is what is technically known as a utility scale. So in the following, the existence of a utility function $u(\omega,a)$ on $\Omega \times A$ is supposed. For each possible combination of an action and a state of nature this function indicates how positively the decision maker evaluates the outcomes. When A is discrete, the utility function will be notated by $u_t(\omega)$.

The decision problem would be easy if the true state of nature were known. If nature is in state ω_0, the best action is the one for which $u(\omega_0,a)$ is maximal. In most decision problems, however, nature does not fully disclose its true state; all we have at hand are fallible data — for example, information from a sample, or subjective beliefs. A neat way of formalizing this is to assume the existence of a random variable Z representing the outcome of some experiment or measurement conducted to get known the true state of nature, and whose distribution depends on ω. In the following it is assumed that the family of probability distributions of Z with distribution functions $F(z) = F_Z(z;\omega)$ is known.

If the true state of nature is not known with certainty, the decisions based on A are likely to be less than optimal. A rational approach, then, is to look for a decision rule that optimizes the outcomes across repetitions of the same

decision problem, but with nature in possibly different states. The main purpose of decision theory is to find such rules and to study their properties. Formally, a (nonrandomized) decision rule is a prescription specifying for each possible value z of Z which action a $\in$ A has to be taken. Hence, a proper notation is to write the decision rule δ as a mapping from the data Z to the action space A: $A = \delta(Z)$. Due to the fact that Z is a random variable, using δ implies that the actions are also random. At first sight this may seem embarrassing. However, there is no way to escape the random nature of our data about ω. Moreover, the decision maker is free to choose whatever rule we would like to have from the set of possible mappings from Z to A, which gives him the opportunity to select an optimal one.

It is obvious that our criterion for selecting an optimal decision rule should have to do with the utilities of the decision outcomes, $u(\omega,a)$. The function $A = \delta(Z)$ implies that these utilities must be considered as realizations of a random variable $U = u(\omega,\delta(Z))$. In such cases it is natural to replace this variable by its expectation. Therefore, we define the expected utility as:

$$(1) \qquad R(\omega,\delta) \equiv E[u(\omega,\delta(Z))] = \int u(\omega,\delta(z))dF(z)$$

If nature is in state ω and decision rule δ is adopted, then (1) shows the utility to be expected. However, the actual

actions still depend on the values taken by Z, so their utility usually will vary about the value of (1).

Bayes rules

It is still not possible to define a criterion for a uniformly best decision rule δ using the expected utility function in (1). The reason is its dependency on the unknown value ω. One way out of this problem would be to make a sensible choice for this value, say ω_0, and to select a δ_0, such that the expected utility $R(\omega_0,\delta)$ is maximal. This approach is taken, for instance, in minimax theory where ω_0 is selected as a value of Ω that represents the least favorable state of nature to the decision maker. As a consequence, the decision maker is guaranteed that the minimum expected utility for δ_0 is never smaller than $R(\omega_0,\delta_0)$.

Another approach is not to select one special state Ω of nature, but to assume a distribution function $G(\omega)$ over Ω. This may represent the decision maker's subjective probabilities of the possible states of nature or its empirical distribution. For both interpretations $G(\omega)$ is known as the a priori distribution (or prior) because it represents knowledge about ω available before the data $Z = z$ are observed. Having $G(\omega)$, we are now able to define the Bayes utility of decision rule δ as

$$(2) \qquad B(\delta) = E[R(\omega,\delta)] = \int R(\omega,\delta)dG(\omega).$$

In the literature this quantity is also known as the Bayes risk of the decision procedure, although, strictly speaking, this name is only proper if a loss function instead of a utility function is used.

The Bayes utility in (2) only depends on the decision rule. It now seems obvious to select from the class of possible rules the one, say δ^*, that maximizes the Bayes utility:

$$(3) \qquad B(\delta^*) = \max_{\delta} B(\delta) .$$

Rules satisfying (3) are known as Bayes rules. Throughout this paper it will be assumed that the quantities used in the above definition of a Bayes rule exist for the problem at hand (though it will not necessarily be true that the problem has a unique Bayes rule).

Monotone Bayes rules

For the actual maximization in (3) it would be helpful if the (possibly infinitively) large set of all possible rules δ could be restricted to some subset of a tractable form.

In addition to this technical consideration, there also is a somewhat less lofty reason for which the attention

sometimes has to be restricted to a subset of possible rules. This has to do with the acceptability of some types of rules among those involved in the decision procedure. In education, for example, students, teachers and administrators are familiar with selection decisions in which the decision rules have a monotone form: students are admitted to a program if their grade-point average or their test score are above a certain cutting point and rejected otherwise. It would mean a shock to all parties if some institutions changed the form of their selection rule and started, e.g., admitting students with low or high scores while rejecting those with intermediate ones. However, the restriction to monotone rules is only correct if they constitute an essentially complete class (Ferguson, 1967, p. 55); otherwise rules with a higher expected utility are wrongly ignored.

The conditions under which a class of monotone rules is essentially complete are known (Chuan, Chen & Novick, 1981; Ferguson, 1967, sect. 6.1). Two conditions have to be met: First, the so-called posterior distribution of ω given $Z = z$ should be stochastically increasing, that is, if $F(\omega|z)$ is the distribution function of ω given $Z = z$, $z_1 \geq z_2$ must imply $F(\omega|z_1) \leq F(\omega|z_2)$ for all ω. Second, there should be an ordering of the actions for which the difference between the utility functions for adjacent pairs of actions changes sign at most once. For the decision problems dealt with in this paper, the conditions will be made more specific below.

A Classification of Test-based Decision Making

The use of test scores for decision making in education and psychology can be classified in a simple way (van der Linden, 1985a, 1985b). In each of these settings three basic elements can be identified and each type of decision making can be viewed as a specific configuration of these elements. In general, four different types of decision making can be distinguished. Further, for each of these types four possible restrictions or refinements can apply. This classification of test-based decision making will now be elaborated. In the next sections it is then shown how Bayesian decision theory can be applied to the problems in this classification.

Basic Elements of Test-based Decisions

Each type of decision making can be identified as a specific configuration of one or more of the following elements:

(1) A test that provides the scores on which the decisions are based;

(2) One or more treatments with respect to which the decisions are made;

(3) One or more criteria by which the successes of the treatments are measured.

The term "test" is used here mainly because of the focus of this paper. It could easily be replaced by any other measuring instrument or source of data without invalidating the content of this paper. Likewise, "treatment" is a generic

term here, referring to whatever manipulation, experiment, or program is used to change the condition of individuals. Examples of treatments in educational and psychological settings are: instructional programs, applications of AV-materials, or psychological therapies. It should be noted that information about the success of the treatment on the criterion may be provided by any source. In this paper it is assumed that the information is quantitative by nature. In practice, the criterion is often measured by another test.

Types of Decisions

Four basic types of decisions are distinguished. Each type of decision can be represented by a unique flowchart containing one or more of the above elements.

(1) Selection decisions. In selection problems the decision is the acceptance or rejection of individuals for a treatment. A typical feature of the selection decision is that the test is administered before the treatment but that the criterion is measured afterwards. Well-known examples of selection decisions are selection of personnel in industry and admission of students to educational programs. Figure 1 gives a flowchart displaying the structure of the selection

---

Insert Figure 1 about here

---

decision. Selection research has had a long tradition in educational and psychological testing in which the selection decision was viewed as a prediction problem. Until Cronbach and Gleser (1965), the usual approach was to establish regression lines or expectancy tables to predict criterion scores and to accept individuals with predicted criterion scores above a given threshold value.

(2) Mastery decisions. Unlike selection decisions, mastery decisions are made after a treatment. The content of the decision is whether or not individuals who followed the treatment are successful on the criterion. A further feature is that the criterion is internal and not external to the test. It is the unreliability of the test as a representation of the criterion that creates the mastery decision problem. Due to measurement error, the possibility of false-negative and false-positive decisions exists, and it is the task of the decision maker to minimize their consequences. Figure 2 shows the formal structure of the mastery decision problem.

Insert Figure 2 about here

Examples of mastery decisions are pass-fail and certification decisions in education.

(3) Placement decisions. In placement problems, several alternative treatments are available and it is the decision

maker's task to assign individuals to the most promising treatment. The same test is administered to each individual and the success of each treatment is measured by the same criterion. Placement decisions differ from selection decisions by the fact that more than one treatment is available and that each individual is assigned to a treatment. The case of a placement decision with two treatments is shown in Figure 3. Examples of placement decisions can be found in individualized instruction where

---

Insert Figure 3 about here

---

students are allowed to follow different routes through instructional units but, regardless of routes, the same criterion is appropriate. Aptitude-treatment interaction (ATI) research has given the main impetus to interest in the placement problem. In ATI research, the traditional approach has been regression analysis with a separate regression line for each treatment assigning individuals to the treatment with the largest predicted criterion score.

(4) Classification decisions. As is clear from Figure 4, the

Insert Figure 4 about here

---

difference between placement and classification decisions is that in the latter, each treatment has its own criterion. Further properties of the two types of decisions are equal. Examples of classification decisions arise in vocational guidance situations where most promising careers or training programs must be identified. The most popular approach to classification decisions has been the use of linear-regression techniques. Each criterion is then mapped on a common utility scale and the decision rule is to assign individuals to the treatment with the largest predicted utility.

Further Restrictions and Generalizations

It should be noted that the above types of decisions are not always met in their pure forms. These decisions often occur in combinations; also, further restrictions or generalizations may apply. An example of a combination of two types of decisions arises in a selection problem where the criterion is unreliably measured. If success on the criterion is defined by a threshold value, then, in fact, after the treatment a mastery decision has tobe taken, and the problem is a selection-mastery decision problem. A combination of a selection and a classification problem is met if more than one treatment is available but not all individuals are accepted for a treatment. In individualized instruction – for instance, as implemented in CAI-systems – decision making can

be viewed as guiding students through a network of several of the above decisions. The simultaneous optimization of such networks will be discussed in the final section of the paper.

For each basic type of decision, one or more of the following restrictions or generalizations may apply:

(1) Multivariate test scores. Instead of a single test, a battery of tests may be used to base the decision on. The use of test batteries has had a long history in the practice of personnel selection and vocational guidance. Formally, the use of test batteries implies decision rules defined on a vector of test scores instead of a single score.

(2) Sequential testing. In a sequential testing strategy, test items are administered until a decision can be made with a desired level of certainty. The recent introduction of the computer in educational and psychological testing has stimulated the interest in sequential testing strategies for decision making. If tests are used in this mode, sequential Bayesian procedures can be used (Lindgren, 1976, sect. 8.5).

(3) Multiple criteria. In some applications, the success of a treatment has to be measured on more than one criterion. Each individual criterion is then supposed to reflect a different aspect of the treatment. Formally, the presence of multiple criteria implies the necessity to define utility functions on a vector of criterion scores instead of a single score.

(4) Multiple populations. The presence of different populations of examinees reacting differently to the test items may create the problem of "fair" decision making. In education,

the problems of fair selection and mastery decisions have been struggled with for a long time, in particular for populations defined by race or sex. Formally, the presence of populations reacting differently to test items implies different probability distributions of test and criterion scores for each population. In addition, the decision maker may have different utilities associated with different populations. As a result, a separate decision rule has to be established for each population.

(5) Quota restrictions. So far it has been assumed that the number of vacancies in each treatment is free. Due to the shortage of resources, however, these numbers may be constrained. Consequently, Bayes rules for quota-restricted decisions have to be found by methods of constrained optimization.

Conclusion

The above classification of test-based decisions shows four basic types of decisions that may occur separately or in combination. Further, for each decision one or more refinements or generalizations may hold. In addition, the utility structure and probability distributions may vary from problem to problem. However, the key to finding Bayes rules for each possible decision is still the optimization of its Bayes utility. This will now be illustrated for a sample of decision problems.

Selection Decisions with Linear Utility

In selection decisions, test scores are used to decide on the acceptance or rejection of individuals for a treatment, with success measured on a future criterion. In order to apply the framework of Bayesian decision theory, the "unknown state of nature" should now be interpreted as the individual's unknown criterion score and the "data" about this state is provided by the test score. For a randomly sampled individual, let the criterion be a continuous random variable Y, with as possible states, success ($Y \geq y_c$) and failure ($Y < y_c$). The test score is assumed to be a discrete random variable X with possible values $x = 0, 1, \ldots, n$ (number-right score). The information in X on Y is given by a joint probability function $k(x,y)$. Since the conditions for a monotone decision are assumed to be met, the Bayes rule for the selection problem is a cut-off score x_c, with an acceptance and a rejection decision for $X \geq x_c$ and $X < x_c$, respectively.

Formally, utility is a function defined on the true state of nature with a possibly different form for each action. A moment's reflection shows that in the present problem utility should be an increasing function of the criterion for the acceptance decision, but a decreasing function for the rejection decision: The higher the criterion score of an accepted individual, the higher the utility of the decision; whereas the opposite holds for a rejected

individual. A linear utility function that meets this property is given in van der Linden and Mellenbergh (1977):

$$(4) \qquad u(y) = \begin{cases} b0(y_c-y) + a0 & \text{for } x < x_c \\ b1(y-y_c) + a1 & \text{for } x \geq x_c, \quad b0, b1 > 0 . \end{cases}$$

This function, which is shown in Figure 5, consists of two additive components:

(1) $b_0(y_c-y)$ and $b_1(y-y_c)$ represent amounts of utility dependent on the difference between the criterion score and success threshold y_c with constants of proportionality b_0 and b_1;

(2) a_0 and a_1 are amounts of utility independent of the criterion score but dependent on the decision. They can be used, for instance, to allow for treatment costs.

Insert Figure 5 about here

When sampling individuals from the population, the Bayes utility for decision rule x_c is equal to

$$(5) \qquad R(x_c) = \sum_{x=0}^{x_c-1} \int [b0(y_c-y) + a0]k(x,y)dy +$$

$$\sum_{x=x_c}^{n} \int [b_0(y-y_c) + a_1]k(x,y)dy.$$

Using $k(x,y) = g(y|x)h(x)$, $\int g(y|x)dy = 1$ and

$\int yg(y|x)dy = E(Y|x)$, it follows

$$(6) \qquad R(x_c) = \sum_{x=0}^{x_c-1} \{b_0[E(Y|x)-y_c] + a_0\}h(x) -$$

$$\sum_{x=x_c}^{n} \{b_1(E(Y|x)-y_c] - a_0\}h(x).$$

where $E(Y|x)$ is the regression function of Y on X.

Completing the first sum

$$(7) \qquad R(x_c) = \sum_{x=0}^{n} \{b_0[E(y|x)-y_c] + a_0\}h(x) -$$

$$\sum_{x=c}^{n} \{(b_0+b_1)[E(Y|x)-y_c] + (a_0-a_1)\}h(x).$$

Since the first sum is now a constant, $b_0 + b_1 > 0$, $h(x) \geq 0$ for all x, and the monotonicity conditions guarantee that $E(Y|x)$ is increasing in x, (7) is maximal for the value of the smallest value of x_c for which

$$(8) \qquad (b_0+b_1)\,[E(Y|x)-y_0] + (a_0+a_1)$$

is not negative. If the monotonicity conditions are not strict or it does not hold that $h(x) > 0$ in the neighborhood of the solution, this value of x may not be unique. Throughout this paper it will be assumed that conditions like these are fulfilled.

The regression function $E(Y|x)$ can easily be estimated by drawing a sample from the population and administering the treatment.

It should be noted that (8) is the difference between the two conditional expected utilities given $X = x$ associated with the acceptance and rejection decision. This is clear from inspection of the bracketed terms in (6) and (7). The expectations are known as posterior expected utilities; they can be considered the expected utilities after the observation $X = x$ has been made. For a monotone decision it holds in general that the optimal cut-off score is located at the point at which the posterior expected utilities cross (e.g., De Groot, 1970, sect. 8.9). In the following, this property will be used without further validation.

A closed-form solution exists if the regression function is linear, that is, if

$$E(Y|x) = \beta x + \alpha \tag{9}$$

Then, it follows from (8) that x_c is the smallest value larger than

$$\frac{y_c - \alpha}{\beta} + \frac{a_0 - a_1}{\beta(b_0+b_1)} . \tag{10}$$

An interesting case arises if $a_0 = a_1$. Under this condition the second term in (10) vanishes and the solution contains the regression parameters only. For this and other properties of the linear utility function, see van der Linden and Mellenbergh (1977).

Multiple Populations

As noted earlier, the presence of multiple populations in selection decisions creates the problem of fair selection if each population reacts differently to the test items. In such a case, the test items are often said to be "biased" against one or more of the populations.

The only thing needed to deal with multiple populations in selection decisions seems to include separate probability distributions of test and criterion scores for each population in the model to allow for differential item

properties. However, the problem of fair selection often involves the notion of "disadvantagedness" as well, in particular when some of the populations are defined by race or sex. As Gross and Su (1975) and Novick and Petersen (1977) argue, this aspect of fair selection is only a question of utilities. A selection rule is "fair" if those involved in the decision process accept the utility structure underlying the decision rule. Hence, in addition to separate probability distributions, separate utility functions are needed to allow for different utility structures for the populations.

Now the above selection model can easily be adapted to the case of multiple populations (Mellenbergh & van der Linden, 1981). Let $i = 1, \ldots, p$ denote the populations in the selection problem. Then, (4) has to be replaced by

$$(11) \qquad u_i(y) = \begin{cases} b_{0i}(y_c - y) + a_{0i} & \text{for } x < x_{ci} \\ b_{1i}(y - y_c) + a_{1i} & \text{for } x \geq x_{ci}, \quad b_{0i}, b_{1i} > 0 \end{cases}$$

and the probability functions $k_i(x,y)$, $i = 1, \ldots, p$, are now allowed to vary across populations. The Bayes utility for a random individual is defined analogous to (5) as $R_i(x_{ci})$. Let S_i be the relative size of population i. Then, when sampling randomly from the total population under consideration, the Bayes utility is equal to

$$R(x_{c1},\ldots,x_{cp}) = \sum_{i=1}^{p} s_i R_i(x_{ci}).$$

But this is just a weighted sum of the Bayes utilities of the separate populations. Hence, the decision procedure is optimal if, analogous to (5) to (8), for each population a separate Bayes rule is derived.

An example of the model with real test data is given in Mellenbergh and van der Linden (1981).

Mastery Decisions with Threshold Utility

In the mastery decision problem, the unknown state of nature is the individual's true score on the criterion variable measured by the test. The data are the observed test scores. The true score is defined as the expected proportion of test items a given individual solves correctly. The true score of a random individual is denoted as T, with possible values $t \in [0,1]$. A mastery level t_c is assumed, and an individual is considered to master the criterion if $t \geq t_c$, and not to master it otherwise. Unlike the selection problem, there is no way whatsoever to measure the criterion variable directly. Hence, a test model is needed to derive the statistical relation of X to the true score T.

For the purpose of illustration, a threshold utility function is chosen. Although Figure 6 shows that this

function has a jump at t_c that may be less realistic in some

---

Insert Figure 6 about here

---

applications, it has been studied extensively in the mastery testing literature (Hambleton & Novick, 1973; Huynh, 1976; Mellenbergh, Koppelaar, & van der Linden, 1977; van der Linden, 1980, 1982; Wilcox, 1977). The threshold utility function is defined by the following four constants

$$(12) \qquad u(t) = \begin{cases} u00 & t < t_c,\ x < x_c \\ u10 & t \geq t_c,\ x < x_c \\ u01 & t < t_c,\ x \geq x_c \\ u11 & t \geq t_c,\ x \geq x_c \end{cases}$$

However, since the derivation following below holds for any positive linear rescaling of (12), it will be assumed for convenience that $u_{00} = u_{11} = 0$.

Let the joint distribution of X and T be given by the probability function $k(x,t)$, and let $g(t|x) \equiv k(x,t)/h(x)$. For the sake of illustration, the optimal cut-off score x_c will now be derived using a comparison of the posterior expected utilities for the mastery and nonmastery decision.

If the mastery decision is taken for a random individual with test score X = x, the posterior expected utility is equal to

$$(13) \qquad E_1(u(t)|x) = \int_0^1 u(t)g(t|x)dt = u_{01} \int_0^{t_c} g(t|x)dt$$

For the nonmastery decision

$$(14) \qquad E_0(U(T)|x) = \int_0^1 u(t)g(t|x)dt = u_{10} \int_{t_c}^1 g(t|x)dt.$$

Suppose x were continuous. Then (13) and (14) would cross at the value of x for which

$$u_{01} \int_0^{t_c} g(t|x)dt = u_{01} \int_{t_c}^1 g(t|x)dt$$

or

$$(15) \qquad \int_0^{t_c} g(t|x)dt = \frac{u_{10}}{u_{01} + u_{10}} .$$

However, X is a discrete test score. Therefore, the Bayes utility is maximal if x_c is chosen to be the smallest integer value larger than the solution to (15).

It should be noted that this solution holds for any test model providing the probability function in the left-hand side of (15), and that it can be calculated only once such a model is specified.

A usual choice in mastery testing is the beta-binomial model (Huynh, 1976; Mellenbergh, Koppelaar, & van der Linden, 1977). In the model it is assumed that (1) the conditional distribution of X given T = t is the binomial,

$$(16) \qquad f(x|t) = \binom{n}{x} t^x (1-t)^{n-x} ,$$

and (2) the marginal distribution of T is the beta distribution with probability density function

$$(17) \qquad b(t) \equiv B^{-1}(v, w-n+1) t^{v-1} (1-t)^{w-n}, \qquad v > 0,\ w > n-1,$$

where $B(v, w-n+1) \equiv \int_0^1 t^v (1-t)^{w-n} dt$ is the complete beta function (e.g., Johnson & Kolz, 1970, chap. 24). The choice of (16) is motivated by the fact that the responses of a fixed individual to a series of test items can often be described as a sequence of Bernoulli trials, whereas (17) defines a flexible family of distributions on [0,1] that contains most true-score distributions occurring in practice. Keats and Lord (1962) have shown that moment estimators of v and w exist that are a simple function of μ_X and the KR-21 reliability coefficient. They also found a satisfactory fit

of the beta-binomial model to test score distributions ranging widely in form.

For integer values of v and w, it holds that

$$(18) \qquad \int_0^{t_c} g(t|x)dt = \int_0^{t_c} f(x|t)b(t)dt$$

$$= \sum_{\gamma=v+x}^{v+w} f(\gamma|t_c)$$

where $f(.|t_c)$ is the binomial probability function with success parameter t_c (Johnson & Kolz, 1970, sect. 24.6). Thus, a suitable estimate of the integral in (15) can be obtained via a table of the cumulative binomial.

A Numerical Example

Using the beta-binomial model, the result in (15) was used to calculate the optimal cut-off score x_c for tests of length $n = 20$, mastery threshold $t_c = 14$, and utility ratio $u_{10}/u_{01} = 1$. The data were simulated such that the average true score μ_T and the KR-21 reliability coefficient varied systematically (van der Linden, 1984). Table 1 shows how x_c depends on these two parameters. For tests of high

---

Insert Table 1 about here

---

reliability x_C varies hardly with μ_T, but this robustness is lost quickly for tests of lower reliability. For KR-21 = .05, $\mu_T \leq 13$ yields $x_C \geq 20$, but the optimal cut-off score drops immediately to below zero for $\mu_T > 13$. Only tests of high reliability yield stable cut-off scores.

An unexpected phenomenon in Table 1 is the opposite direction in which x_C varies with μ_T. If the average true score μ_T goes down, x_C goes up. Hence, for low performing populations the cut-off score should be set high, whereas it should be low for high performers. At first sight, this goes against our intuition. However, it is a logical consequence of our criterion of maximal expected utility. An analysis of (12) to (15) reveals the following: For high performing populations, almost all individuals are above the mastery threshold t_C. Therefore, the Bayes utility of the decision procedure tends to consist only of contributions from true masters and false nonmasters. Since the proportion of true masters among these two categories depends on the cut-off score x_C, and the utility of a true master typically is larger than the utility of a false nonmaster, x_C will take a low value. This phenomenon was dubbed "the regression-from-the mean effect" in van der Linden (1980). It is a reversion of the well-known regression-to-the-mean effect due to the fact that inthe derivation of (15) the conclusion goes from the true score t_C to x_C instead of the other way around.

Placement Decisions with Normal–Ogive Utility

The typical feature of the placement decision is the presence of more than one treatment and the fact that each individual is assigned to a treatment. As before, the unknown state of nature is the individual's criterion score after a treatment. Hence, although there is one criterion common to all treatments, each individual has a different unknown state of nature for each criterion.

Let Y be the criterion common to the treatments $j = 1, 2, \ldots, m$, and let X denote the test score again. It is assumed that the treatments are ordered by the strict monotonicity conditions for the placement decision given in van der Linden (1981), so that an optimal cut–off score x_j exists for the decision between treatments j and $j + 1$. For the sake of illustration, the normal–ogive utility function introduced by Novick and Lindley (1978; see also Berhold, 1973) will be used (see also Figure 7):

$$(19) \qquad u_j(y) = \Phi\left(\frac{y-\mu_j}{\sigma_j}\right),$$

where Φ is the normal distribution function with parameters μ_j and σ_j.

Insert Figure 7 about here

The relevant quantity is the posterior expected utility for treatment j after test score $X = x$. This will be denoted as $E_j(u_j(Y)|x)$. It is assumed that the conditional distribution of Y given $X = x$ is normal with conditional mean and variance $E_j(Y|x)$ and $Var_j(Y|x)$, respectively. It follows that

$$(20) \qquad E_j(u_j(Y)|x) = \Phi\left\{ \frac{E_j(Y|x) - \mu_j)}{[Var_j(Y|x) + \sigma^2_j]^{1/2}} \right\}$$

The optimal cut-off score x_j is now the smallest value of x for which the posterior expected utility for treatment $j + 1$ is larger than for treatment j, that is, the smallest value of x for which

$$(21) \qquad \Phi\left\{ \frac{E_{j+1}(Y|x) - \mu_{j+1}}{[Var_{j+1}(Y|x) + \sigma^2_{j+1}]^{1/2}} \right\} - \Phi\left\{ \frac{E_j(Y|x) - \mu_j}{[Var_j(Y|x) + \sigma^2_j]^{1/2}} \right\}$$

is positive. But this is also the smallest value of x for which

$$(22)\qquad \frac{E_{j+1}(Y|x) - \mu_{j+1}}{[Var_{j+1}(Y|x) + \sigma^2_{j+1}]^{1/2}} - \frac{E_j(Y|x) - \mu_j}{(Var_j(Y|x) + \sigma^2_j]^{1/2}}$$

is positive. If it is known how the posterior expectations and variances depend on x, the optimal cut-off score can be found graphically or numerically from (22).

Suppose now that the posterior expectations are linear and that homoscedasticity may be assumed. That is,

$$(23)\qquad E_j(Y|x) = \beta_j x + \alpha_j, \qquad j = 1, \ldots, m$$

and

$$(24)\qquad Var_j(Y|x) = Var_j(Y.X) \qquad \begin{matrix} x = 0, \ldots, n \\ j = 1, \ldots, m \end{matrix}$$

where $Var_j(Y.X) = (1-[Cor_j(Y,X)]^2)Var_j(Y)$ and $Cor_j(Y,X)$ is the linear correlation coefficient between Y and X for j. Substituting (23) and (24) into (22) yields as a result that x_j is the smallest value larger than

$$(25)\qquad [\varepsilon_j(\mu_{j+1}-\alpha_{j+1}) - \varepsilon_{j+1}(\mu_j+\alpha_j)]/(\varepsilon_j\beta_{j+1}-\varepsilon_{j+1}\beta_j)$$

with

$$\varepsilon_j = [Var_j(Y.X) + \sigma_j^2]^{1/2}$$

More theory on placement decisions is given in van der Linden (1981).

Quota-Restricted Placement

In the above placement model, it was assumed that the number of vacant places in the treatments was unrestricted. In practice, however, quota restrictions regularly apply, and then a modification of the decision rule is necessary.

Following Chuang, Chen and Novick (1981), three kinds of quota restrictions are distinguished:

(1) Exactly N_j individuals should be assigned to treatment $j = 1, \ldots, m$, where $\sum_{j=1}^{n} N_j = N$ (number of examinees);

(2) At least N_j individuals should be assigned to treatment $j = 1, \ldots, m$, where $\sum_{j=1}^{n} N_j \leq N$;

(3) At most N_j individuals should be assigned to treatment $j = 1, \ldots, m$, where $\sum_{j=1}^{n} N_j \leq N$.

In a (strictly) monotone placement problem, the posterior expected utilites of adjacent pairs of treatments cross at most once in the range of test scores. Therefore, if the N examinees are arranged in decreasing order of test scores, the following placement rules maximize the Bayes utility in the above three cases:

(1) Beginning with the highest scoring examinee, N_m places in treatment m are filled, then N_{m-1} places in treatment m-1, and so on.

(2) Suppose the (k−1)th examinee in the order of examinees has been assigned to treatment j. Then the following rule for k is optimal:

(i) If $x_k \geq x_{j-1}$, assign k to treatment j.

(ii) If $x_k < x_{j-1}$ and treatment j has not received N_j examinees, then choose from the following rules the one with the larger posterior expected utility: (a) assign k to j; (b) assign k to j−1, reassigning the lowest scoring examinee in j+1 to j (and, if j+1 then has fewer than N_{j+1} examinees, the lowest scoring one in j+2 to j+1; etc.).

(iii) If $x_k < x_{j-1}$ and j has received N_j examinees, then assign k to j−1.

(3) Again suppose the (k−1)th examinee has been assigned to treatment j.

(i) If $x_k < x_{j-1}$, then assign k to the treatment in {j−1, j−2,..., 1} with the largest posterior expected utility.

(ii) If $x_k \geq x_{j-1}$ and treatment j has received N_j examinees, then choose from the following rules the one with the larger posterior expected utility: (a) assign k to j−1; (b) assign k to j reassigning the highest scoring examinee in j to j+1 (and, if j+1 had already N_{j+1} examinees, the highest scoring one in j+1 to j+2; etc.).

(iii) If $x_k \geq x_{j-1}$ and j has not received N_j examinees, then assign k to j.

Classification Decisions with Threshold Utility

The same notation as for the problem of the placement decision will be used. However, because every treatment now has its own criterion, the criterion variable is treatment dependent and will be denoted as Y_j.

Suppose that for each criterion a success threshold d_j can be defined. Then the following threshold utility function may be a proper choice:

$$(26) \qquad u_j(y_j) = \begin{cases} w_j & \text{for } y_i \geq d_j \\ v_j & \text{for } y_j < d_j \end{cases}$$

with

$$w_j > v_j \qquad\qquad j = 1, \ldots, m$$

The last condition simply states that the utility of a success on treatment j is larger than the one of a failure.

Let $\Omega_j(d_j|x) = \int^{d_j} g_j(y_j|x)dy_j$, where $g_j(y_j|x)$ is the probability density function of Y_j given $X = x$ and treatment j. This quantity defines the probability of a failure on criterion Y_j after treatment j. Van der Linden (1987) shows

that the following conditions are sufficient for a monotone Bayes rule in a classification problem with threshold utility:

(27) $w_{j-1} - v_{j-1} \leq w_j - v_j$ $\quad j = 2, \ldots, m$

(28) $\Omega_j(d_j|x)$ is decreasing in x $\quad j = 1, \ldots, m$

(29) $\Omega_{j-1}(d_{j-1}|x) - \Omega_j(d_j|x)$ is increasing in x $\quad j = 2, \ldots, m$

The condition in (27) indicates that the relevant order of the treatments is with respect to $w_j - v_j$. The other two conditions state that the probability of a failure should be decreasing in the test score, but that the difference between these probabilities for two adjacent treatments should increase.

The optimal cut-off score for a decision between treatments j and j+1 is found by comparing posterior expected utilities. If an individual with test score X = x is assigned to treatment j, the posterior expected utility is equal to

(30) $E_j(u(Y_j)|x) = \Omega_j(d_j|x)v_j + [1-\Omega_j(d_j|x)]w_j.$

The interest is in the value of x at which $E_j(u_j(Y_j)|x)$ and $E_{j+1}(u_{j+1})|x)$ cross, that is, the value of x for which

$$\Omega_j(d_j|x)v_j + [1-\Omega_j(d_j|x)]w_j$$

$$= \Omega_{j+1}(d_{j+1}|x)v_{j+1} + [1-\Omega_{j+1}(d_{j+1}|x)]w_{j+1}$$

or

$$(31) \quad (w_j-v_j)\Omega_j(d_j|x) - (w_{j+1}-v_{j+1})\Omega_{j+1}(d_{j+1}|x) + w_{j+1} - w_j = 0$$

However, since X is discrete, the optimal cut-off score is the smallest value of x for which the left-hand side of (31) is positive. The conditions in (27) to (29) guarantee that this expression is an increasing function of x. If it takes the value of zero outside the range of possible test scores, the cut-off score is set at its corresponding border.

An interesting case arises if the utility parameters w_j and v_j do not vary across treatments. Then it follows that $w_j - v_j = w_{j+1} - v_{j+1}$ and $w_{j+1} - w_j = 0$, and the left-hand side of (31) reduces to

$$(32) \quad \Omega_j(d_j|x) - \Omega_{j+1}(d_{j+1}|x) \ .$$

This expression reminds us of (21). Analogous to the argument following (21), it can be shown that for the choice of a normal distribution function for $\Omega_j(.|x)$ together with linear regression functions $E_j(Y|x) = \beta_j x + \alpha_j$ and homosce-

dasticity, an optimal value of x_j equal to the smallest value of x larger than

$$(33) \qquad \frac{(d_{j+1}-\alpha_{j+1})[\mathrm{Var}_j(Y.X)]^{\frac{1}{2}} - (d_j-\alpha_j)[\mathrm{Var}_{j+1}(Y.X)]^{\frac{1}{2}}}{\beta_{j+1}[\mathrm{Var}_j(Y.X)]^{\frac{1}{2}} - \beta_j[\mathrm{Var}_{j+1}(Y.X)]^{\frac{1}{2}}}$$

is obtained (van der Linden, 1987). This quantity can easily be calculated if estimates of the regression parameters and the pooled variances of Y given X = x are available.

Multivariate Test Scores

The use of a test battery instead of a single test was mentioned earlier as a possible refinement of test-based decision making. How Bayes rules can be derived in such cases will now be shown for the problem of placement decisions.

It is assumed that the preceding theory of placement decisions with threshold utility holds. However, in addition to test score X, a second test score Z with information about the criterion variables Y_j (j = 1, ..., m) is given. Z is also assumed to be defined by number-right scoring and may take the values 0, 1, ..., r. Let $k_j(x,y_j,z)$ be the joint probability function of test scores X and Z and criterion score Y_j, while p(z) is the probability function of Z. Further, let S_j denote the set of ordered pairs (x,z) for which treatment j is assigned. The Bayes utility when sampling from the population is defined as

$$(34) \qquad R(S_1,\ldots,S_m) = \sum_{j=1}^{m} \sum_{S_j} \int u_j(y_j)k_j(x,y_j,z)dy_j$$

It follows immediately that

$$(35) \qquad R(S_1,\ldots,S_m) = \sum_{j=1}^{m} \sum_{S_j} p(z) \int u_j(y_j)q(x,y_j|z)dy_j,$$

where $q(x,y_j|z)$ is the probability function of (X,Y_j) given $Z = z$. Suppose now that (28) and (29) hold for each value of z. Then, for each possible z-coordinate, (35) can be maximized following (30) and (31) with respect to the x-coordinates in S_j. Thus, the optimal sets S_j are defined by a series of (r+1)(m-1) cut-off scores, namely m-1 scores for each of the r+1 possible values of z.

By symmetry, it holds that the z-coordinates of the optimal sets S_j can also be defined by a series of cut-off scores $z_1, \ldots, z_{m-1}$ if the failure probabilities $\Omega_j(dj|z)$ have the properties in (28) and (29) for each value of x. In the testing literature, classification rules that are monotone in two distinct test scores are known as "conjunctive rules". For classification into a given treatment such rules require that the examinee passes a certain cut-off score on x as well as on y. In other words,

it is impossible to compensate a failure on one test by a high score on the other.

Combination of Basic Decisions

As noted before, a well-known example of combinations of basic decisions in education is an individualized instruction system. Figure 8 shows a simple system in which a selection

---

Insert Figure 8 about here

---

decision is followed by a module with a mastery decision after which a placement decision guides the students through two possible sequences of mastery decisions. Real-life systems often have a more involved structure.

In systems of more than one decision point, it is possible to optimize decision rules simultaneously. In doing so, more efficient use of the data in the system can be made. Also, more realistic utility structures can be used.

In order to illustrate the procedure, the simple case of a system with one placement and one mastery decision (van der Linden & Vos, 1986; Vos & van der Linden, 1987) will be dealt with (see Figure 9). The placement test score is denoted as X

Insert Figure 9 about here

($x = 0, 1, \ldots, m$), the observed mastery test score as Y ($y = 0, 1, \ldots, n$), and the true mastery test score as T ($t \in [0,1]$). Further, the following definitions are made: x_c is a cut-off score on x, y_c is a cut-off score on y, t_c is the mastery threshold on t, $j = 1, 2$ are the two possible treatments and $i = 0, 1$ are the possible states of nonmastery ($T < t_c$) and mastery $T \geq t_c$), respectively. Without bothering about conditions for optimality, a monotone rule with maximum utility is looked for; that is, from the pairs of cut-off scores (x_c, y_c), the one with the largest Bayes utility will be chosen. The following threshold utility function is adopted:

$$(36) \qquad u_{ij}(t) = \begin{cases} v00 + wj & \text{for } t < d,\ y < c \\ v10 + wj & \text{for } t \geq d,\ y < c \\ v01 + wj & \text{for } t < d,\ y \geq c \\ v11 + wj & \text{for } t \geq d,\ y \geq c, \quad j = 1, 2. \end{cases}$$

The parameter v is dependent both on the mastery state and the mastery decision, but it is independent of the treatment. In addition, the treatment-dependent parameter w can be used to allow for differences in, for instance, treatment costs. Although this function could be made more realistic by

replacing v by continuous functions of the true score t, it nicely demonstrates how in a simultaneous approach a utility function defined on the ultimate criterion of the system (mastery-nonmastery state) can be brought into a previous decision (placement decision).

The Bayes utility for assignment to treatment 1 and a correct mastery decision is equal to

$$(37) \qquad \sum_{x=0}^{x_c-1} \sum_{y=0}^{y_c-1} \int_0^{t_c} (v_{00}+w_1)\varphi(x,y,t)dt,$$

where $\varphi(x,y,t)$ is the probability function of (X,Y,T). Combining the two possible treatments, two possible mastery decisions and two possible mastery states, the total expected utility when sampling from the population of individuals is equal to the sum of eight expressions like (37). From van der Linden and Vos (1986) it can be verified that this sum can be reduced to the following posterior form:

$$(38) \qquad R(x_c,y_c) = \text{constant} + \sum_{y=y_c}^{n} [(-v_{00}-v_{11}) \int_0^{t_c} \varphi_0(t|y)+v_{11}]\kappa_0(y)$$

$$+ \sum_{x=x_c}^{m} \left\{ v_{00}[\int_0^{t_c} (\rho_1(t|x)-\rho_0(t|x)\right\} dt + w_1 - w_0)\lambda(x)$$

$$+ \sum_{x=x_c}^{m} \sum_{y=y_c}^{n} \Big\{ [v_{00}+v_{11}) \int_0^{t_c} \pi_0(t|x,y)dt - v_{11}]\eta_0(x,y)$$

$$- [(v_{00} + v_{11}) \int_0^{t_c} \pi_1(t|x,y)dt - v_{11}]\eta_1(x,y)]\Big\},$$

with $\varphi_0(t|y)$, $\rho_j(t|x)$, $\pi_j(t|x,y)$ being the posterior probability density functions of T given Y=y, X=x and (Y=y,X=x), respectively, and where $\eta_j(x,y)$, $\lambda(y)$ and $\kappa_j(y)$ are the marginal probability functions.

To obtain the optimal values of x_c and y_c, (38) has to be evaluated numerically for all possible pairs of values, whereupon the pair for which (38) is maximal can be determined. This can be done as soon as accurate estimates for the three different posterior probabilities of a failure in (38) are available. To obtain these estimates, a test theory model is needed.

Conclusion

A review of the applications of Bayesian decision theory to test-based decision making was given. Four basic types of decisions were distinguished, each of which may be subject to further restrictions or generalizations. Decisions may also be made separately or in combination. The test theory literature contains many more results for selection and mastery decisions than was presented in this paper. The

placement and classification decisions as well as some of the possible refinements and generalizations and the case of combinations of decisions have been largely unexplored. Of paramount importance, however, is research on realistic utility functions for test-based decisions. Such functions should not only yields robust decision rules (Vijn & Molenaar, 1981), but also be supported by empirical evidence of their appropriateness (Vrijhof, Mellenbergh & van den Brink, 1983). Developments in this field will be decisive for the applicability of Bayesian theory to test-based decision making.

References

Berhold, M.H. (1973). The use of distribution functions to represent utility functions. Management Science, 19, 825-829.

Chuang, D.T., Chen, J.J., & Novick, M.R. (1981). Theory and practice for the use of cut-scores for personnel decisions. Journal of Educational Statistics, 6, 129-152.

Cronbach, L.J., & Gleser, G.C. (1965). Psychological tests and personnel decisions (2nd ed.). Urbana, Ill.: University of Illinois Press, 1965.

DeGroot, M.H. (1970). Optimal statistical decisions. New York: McGraw-Hill.

Ferguson, T.S. (1967). Mathematical statistics: A decision theoretic approach. New York: Academic Press.

Gross, A.L., & Su, W.H. (1975). Defining a "fair" or "unbiased" selection model: A question of utilities. Journal of Applied Psychology, 60, 345-351.

Hambleton, R.K., & Novick, M.R. (1973). Toward an integration of theory and method for criterion- referenced tests. Journal of Educational Measurement, 10, 159-170.

Huynh, H. (1976). Statistical considerations of mastery scores. Psychometrika, 41, 65-79.

Johnson, N.L., & Kotz, S.S. (1970). Continuous univariate distributions (Vol. 2). Boston: Houghton Mifflin.

Keats, J.A., & Lord, F.M. (1962). A theoretical distribution for mental test scores. Psychometrika, 27, 59-72.

Lindgren, B.W. (1976). Statistical theory (Third Edition). New York: Macmillan Publishing Co., Inc.

Mellenbergh, G.J., Koppelaar, H., & van der Linden, W.J. (1977). Dichotomous decisions based on dichotomously scores items: A case study. Statistica Neerlandica, 31, 161-169.

Mellenbergh, G.J., & van der Linden, W.J. (1981). The linear utility model for optimal selection. Psychometrika, 46, 283-293.

Novick, M.R., & Lindley, D.V. (1978). The use of more realistic utility functions in educational applications. Journal of Educational Measurement, 15, 181-191.

Novick, M.R., & Petersen, N.S. (1976). Towards equalizing educational and employment opportunity. Journal of Educational Measurement, 13, 77-88.

Taylor, H.C., & Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. Journal of Applied Psychology, 23, 565-578.

van der Linden, W.J. (1980). Decision models for use with criterion-referenced tests. Applied Psychological Measurement, 4, 469-492.

van der Linden, W.J. (1981). Using aptitude measurements for the optimal assignment of subjects to treatments with and without mastery score. Psychometrika, 46, 257- 274.

van der Linden, W J. (1982). Passing score and length of a mastery test. Evaluation in Education, 5, 149–165.

van der Linden, W.J. (1984). Over absolute en nog relatievere zak–slaag beslissingen [On absolute and more relative pass–fail decisions]. Tijdschrift voor Onderwijs–research, 9, 243–252.

van der Linden, W.J. (1985a). Decision theory in educational research and testing. In T. Husen & T.N. Postlethwaite, International encyclopedia of education: Research and studies. Oxford: Pergamon Press.

van der Linden, w.J. (1985b). Een overzicht van de moderne testtheorie [A review of modern test theory]. Tijdschrift voor de Psychologie, 40, 380–389.

van der Linden, W.J. (1987). The use of test scores for classification decisions with threshold utility. Journal of Educational Statistics, 12, 62–75.

van der Linden, W.J., & Mellenbergh, G.J. (1977). Optimal cutting scores using a linear loss function. Applied Psychological Measurement, 1, 593–599.

van der Linden, W.J., & Vos, H.J. (1985). Optimale regels voor toetsgebruik in individuele studiesystemen [Optimal rules for test use in individualized study systems].s In W.J. van der Linden (Ed.), Moderne methoden voor toetsconstructie en –gebruik. Lisse, The Netherlands: Swets & Zeitlinger.

van Neumann, J., & Morgenstern, O. (1953). Theory of games and economic behavior. Princeton, N.J.: Princeton University Press.

Vijn, P., & Molenaar, I.W. (1981). Robustness regions for dichotomous decisions. Journal of Educational Statistics, 6, 205-235.

Vos, H.J., & van der Linden, W.J. (1987). Designing optimal rules for instructional decision making in CAI systems. In J. Moonen & T. Plomp (Eds.), Developments in educational software and courseware. Oxford: Pergamon Press.

Vrijhof, B.J., Mellenbergh, G.J., & van den Brink, W.P. (1983). Assessing and studying utility functions in psychometric decision theory. Applied Psychological Measurement, 7, 341-357.

Wald, A. (1950). Statistical decision functions. New York: Wiley.

Wilcox, R.R. (1977). Estimating the likelihood of false-positive and false-negative decisions in mastery testing: An empirical Bayes approach. Journal of Educational Statistics, 2, 289-307.

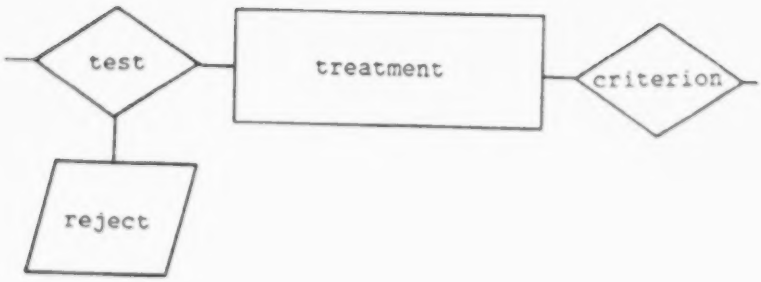

Figure 1. Flowchart of a selection decision

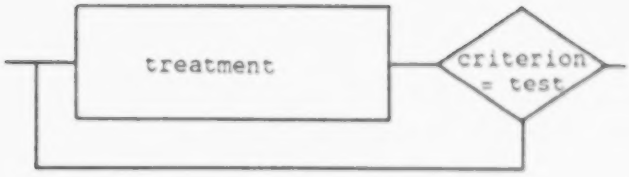

Figure 2. Flowchart of a mastery decision

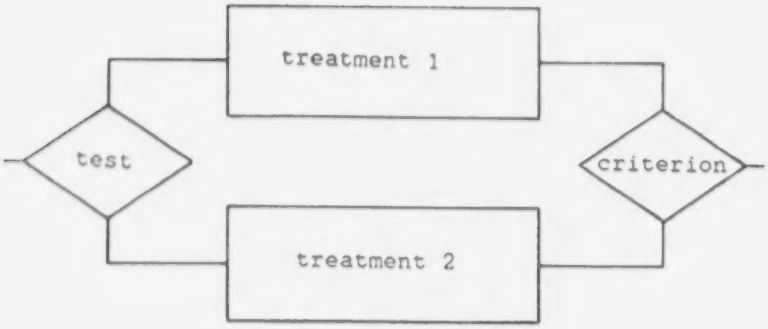

Figure 3. Flowchart of a placement decision (case of two treatments)

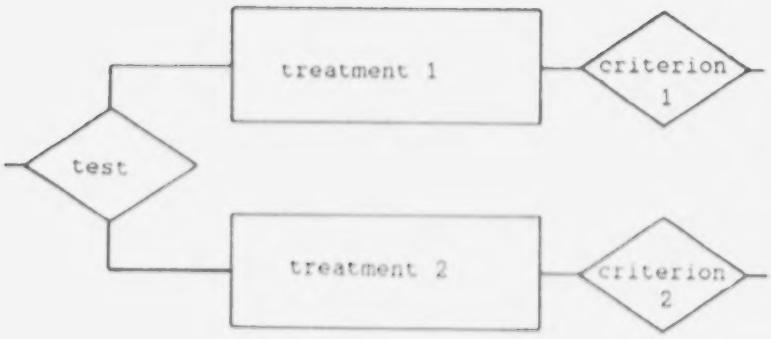

Figure 4. Flowchart of a classification decision (case of two treatments)

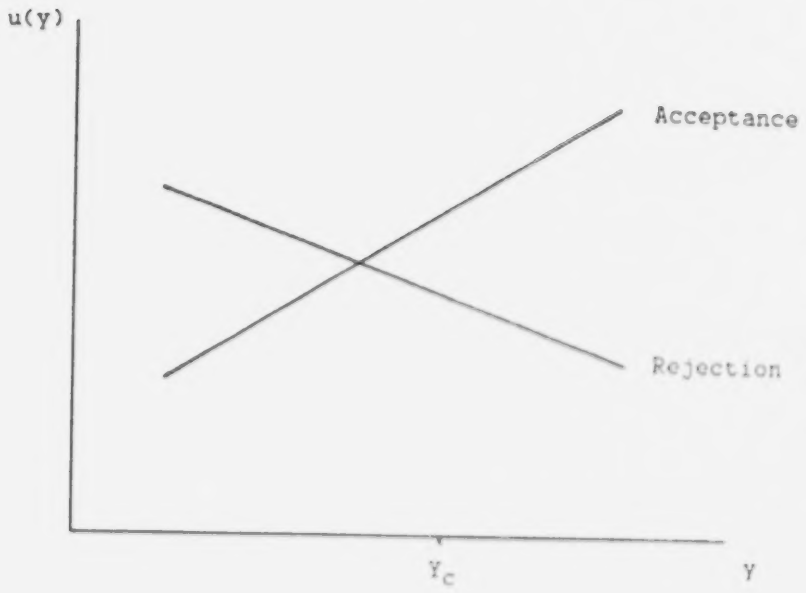

Figure 5. Linear utility function for a selection decision

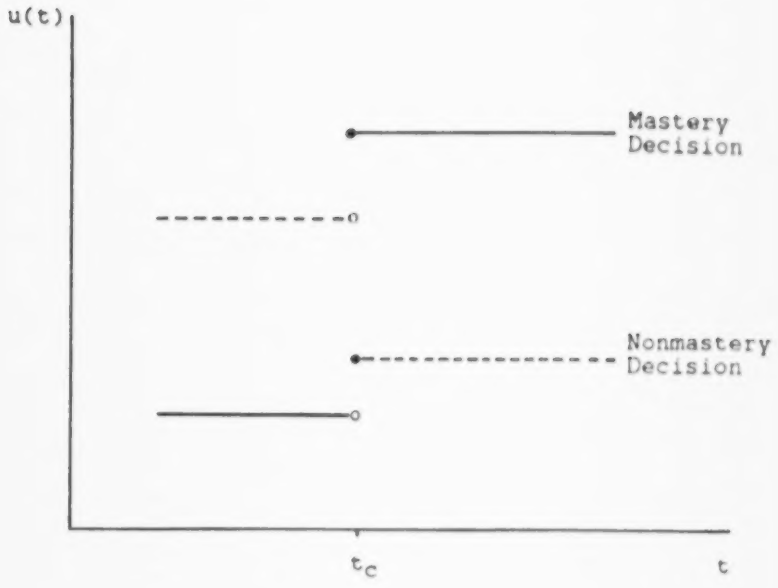

Figure 6. Threshold utility function for a mastery decision

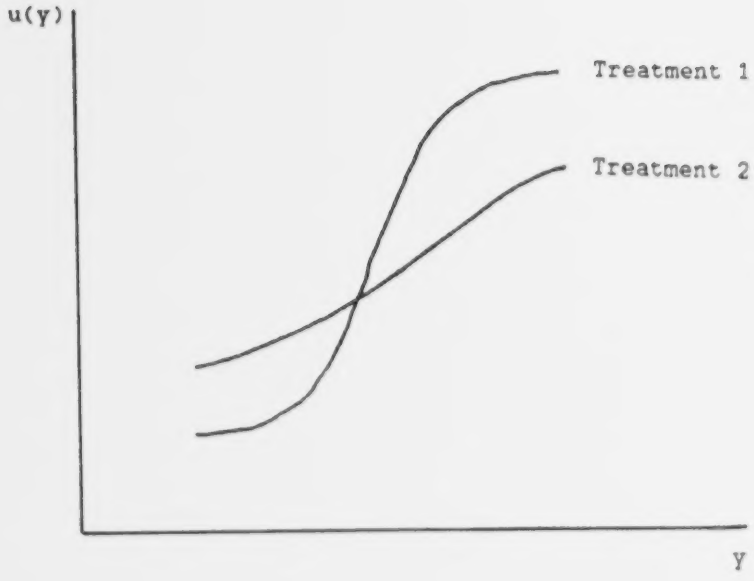

Figure 7. Normal-ogive utility function for a placement decision

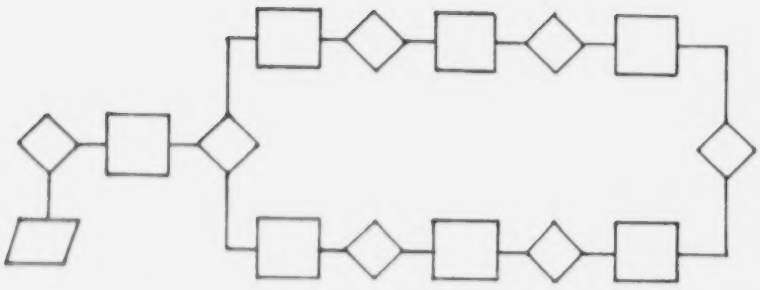

Figure 8. Example of an individualized study system

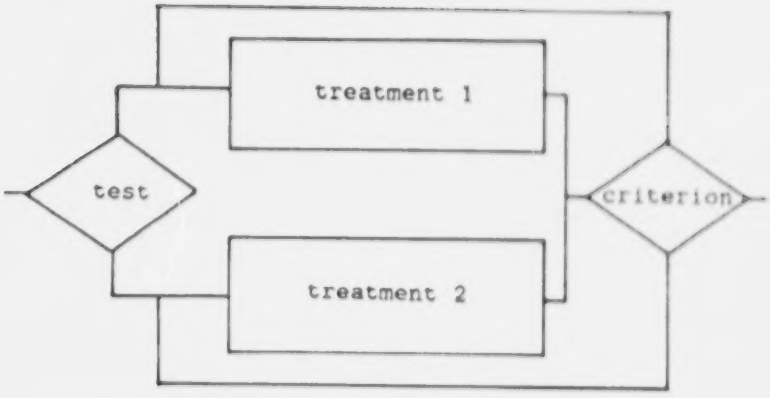

Figure 9. A system of one placement and one mastery decision

Table 1

Optimal Cut-off Scores for Populations Varying in Average True Score and Tests Varying in Reliability ($n=20$, $t_c=14$, $u_{10}/u_{01}=1$)

KR-21	Average True Score									
	1	3	5	7	9	11	13	15	17	19
.95	16	15	15	15	15	15	14	14	14	14
.80	F	19	17	16	16	15	15	14	14	14
.65	F	F	F	F	17	16	15	14	13	12
.50	F	F	F	19	19	17	16	13	12	P
.35	F	F	F	F	F	20	16	13	P	P
.20	F	F	F	F	F	F	19	11	P	P
.05	F	F	F	F	F	F	20	P	P	P

Note F indicates a cut-off score larger than 20 (all students fail); P indicates a cut-off score lower than 0 (all students pass).

Titles of Recent Research Reports

RR-86-1 W.J. van der Linden. The use of test scores for classification decisions with threshold utility

RR-86-2 H. Kelderman, Item bias detection using the loglinear Rasch model: Observed and unobserved subgroups

RR-86-3 E. Boekkooi-Timminga, Simultaneous test construction by zero-one programming

RR-86-4 W.J. van der Linden, & E. Boekkooi-Timminga, A zero-one programming approach to Gulliksen's matched random subtests method

RR-86-5 E. van der Burg, J. de Leeuw, & R. Verdegaal, Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features

RR-86-6 W.J. van der Linden, & T.J.H.M. Eggen, An empirical Bayes approach to item banking

RR-86-7 E. Boekkooi-Timminga, Algorithms for the construction of parallel tests by zero-one programming

RR-86-8 T.J.H.M. Eggen, & W.J. van der Linden, The use of models for paired comparisons with ties

RR-86-9 H. Kelderman, Common item equation using the log-linear Rasch model

RR-86-10 W.J. van der Linden, & M.A. Zwarts, Some procedures for computerized ability testing

RR-87-1 R. Engelen, Semiparametric estimation in the Rasch model

RR-87-2 W.J. van der Linden (Ed.), IRT-based test construction

RR-87-3 R. Engelen, P. Thommassen, & W. Vervaat, Ignatov's theorem: A new and short proof

RR-87-4 E. van der Burg, & J. de Leeuw, Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis

RR-87-5 H. Kelderman, Estimating a quasi-loglinear models for the Rasch table if the number of items is large

RR-87-6 R. Engelen, A review of different estimation procedures in the Rasch model

RR-87-7 D.L. Knol & J.M.F. ten Berge, Least-squares approximation of an improper by a proper correlation matrix using a semi-infinte convex program

RR-87-8 E. van der Burg & J. de Leeuw, Nonlinear canonical correlation analysis with k sets of variables

RR-87-9 W.J. van der Linden, Applications of decision theory to test-based decision making

Research Reports can be obtained at costs from Mediatheek, Department of Education, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.

department of
EDUCATION

A publication by
the Department of Education
of the University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands